

AD-A263 662



28809.11-MA

2

ON THE VALUE OF FUNCTION EVALUATION LOCATION
INFORMATION IN MONTE CARLO SIMULATION

by

Thomas J. DiCiccio and Peter W. Glynn

TECHNICAL REPORT No. 92-17

NOVEMBER 1992

DTIC
ELECTE
MAY 07 1993
S E D

Prepared under the Auspices
of
U.S. Army Research Contract
DAAL-03-91-G-0101

DTIC QUALITY INSPECTED 8

Approved for public release: distribution unlimited.

Reproduction in whole or in part is permitted for any
purpose of the United States Government

93-09501



1198

DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CA 94305

Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

1. Introduction

Monte Carlo sampling strategies are commonly used to numerically evaluate high-dimensional multiple integrals. As is well known, the associated estimators enjoy the nice property that their convergence rates are relatively insensitive to dimensional effects. In particular, the central limit theorem (CLT) implies that Monte Carlo estimators typically converge at rate $n^{-1/2}$, where n is the number of function evaluations made, and this rate is independent of the dimension d of the integral. In contrast, conventional deterministic integration schemes often suffer from a convergence rate of the order $n^{-1/d}$. (See, for example, Davis and Rabinowitz 1984).

Despite the high-dimensional advantages associated with Monte Carlo sampling, it is sometimes possible to improve upon its convergence characteristics. We note that Monte Carlo sampling schemes typically construct estimators that involve sample means. A sample mean is an average over the associated observations in which each observation is equally weighted. (In the Monte Carlo integration setting, an observation is basically just a function evaluation computed at some randomly chosen point.) Such sample means use only the information that is present in the function evaluations themselves. All information about the random locations at which the function evaluations were computed is discarded.

This note concerns the improvements in estimation made possible by taking location information into account. Our main result shows that variance is dramatically reduced for one-dimensional integration problems when location information is exploited. In particular, the convergence rate of the resulting estimator can be improved from rate $n^{-1/2}$ to rate n^{-2} (under suitable smoothness conditions on the integrand). This result suggests that it may be advantageous, in future research, to consider high-dimensional Monte Carlo sampling algorithms that are somehow able to take advantage of the presence of the additional function evaluation location information.

2. Description of the main result

Suppose that we wish to integrate a function f over the unit interval $[0, 1]$. (Note that by a suitable change of variable, all one-dimensional integration problems may be reduced to this form.) The naive Monte Carlo approach to the estimation of

$$\alpha = \int_0^1 f(x) dx$$

begins with generating n independent and identically distributed (iid) uniform $[0, 1]$ random variables U_1, \dots, U_n . The function f is then evaluated at each of the n locations U_1, \dots, U_n , so that $f(U_1), \dots, f(U_n)$ are obtained, and the integral α is estimated by the sample mean

$$\alpha_n = \frac{1}{n} \sum_{i=1}^n f(U_i).$$

Note that α_n is a function purely of $f(U_1), \dots, f(U_n)$, whereas the information available to the simulator comprises the larger set $\{(U_i, f(U_i)) : i = 1, \dots, n\}$. In any case, the CLT implies that if $\sigma^2 = \text{var } f(U_1) < \infty$, then

$$n^{1/2}(\alpha_n - \alpha) \Rightarrow \sigma N(0, 1) \quad (2.1)$$

as $n \rightarrow \infty$, where $N(0, 1)$ denotes a normal random variable with zero mean and unit variance. The weak convergence result (2.1) establishes that the convergence rate of α_n to α is of the order of $n^{-1/2}$ in the number n of function evaluations computed. This convergence rate is relatively slow, especially compared in the one-dimensional setting to the faster convergence rates of a number of deterministic quadrature formulae that are described, for example, by Davis and Rabinowitz (1984).

Because of the random sampling that is being used, the "spacing" between the locations U_1, \dots, U_n will not be regular. In fact, it is well known (see Devroye 1982) that if K_n is the size of the largest subinterval into which $[0, 1]$ is partitioned by U_1, \dots, U_n , then

$$K_n \sim \frac{\log n}{n} \quad \text{a.s.}$$

as $n \rightarrow \infty$. (For sequences $\{a_n\}$ and $\{b_n\}$, the notation $a_n \sim b_n$ is used if $a_n/b_n \rightarrow 1$ as $n \rightarrow \infty$.) Although it may be thought that the slow rate of convergence for α_n is a consequence of the irregular spacing, arguments given below show that this is not the case.

Given that the integral α can be computed via Riemann sum approximations, it makes sense to weight a given function evaluation according to the size of the subinterval over which it is assumed to approximate the original function f . More specifically, let $U_1(n), \dots, U_n(n)$ be the order statistics of the sample U_1, \dots, U_n , so that $U_k(n)$ denotes the k th largest observation in the sample of n iid uniform random variables. We then approximate f over the interval $[(U_{i-1}(n) + U_i(n))/2, (U_i(n) + U_{i+1}(n))/2]$ by the constant value $f(U_i(n))$, to arrive at the approximation

$$\hat{\alpha}_n = \sum_{i=1}^n f(U_i(n))(L_i(n) - L_{i-1}(n))$$

to the integral α , where $L_0(n) = 0$, $L_n(n) = 1$, and $L_i(n) = (U_i(n) + U_{i+1}(n))/2$ for $1 \leq i \leq n-1$. Since the random variable $L_i(n)$ lies at the midpoint between $U_i(n)$ and $U_{i+1}(n)$ for $1 \leq i \leq n-1$, $\hat{\alpha}_n$ is a randomized version of the midpoint quadrature rule of classical numerical integration.

The main result is the following theorem, which is proved in the Appendix.

Theorem 1. Suppose that f is twice continuously differentiable over $[0, 1]$. Then

$$n^2(\hat{\alpha}_n - \alpha) \Rightarrow \frac{f^{(1)}(1)}{2}(\chi_1^2 - 1) - \frac{f^{(1)}(0)}{2}(\chi_0^2 - 1) \quad (2.2)$$

as $n \rightarrow \infty$, where χ_0^2 and χ_1^2 are independent chi-squared random variables each with two degrees of freedom.

Theorem 1 shows that when function evaluation location information is exploited, the rate of convergence can be dramatically improved. Specifically, upon comparison of (2.2) to (2.1), it is evident that using $\hat{\alpha}_n$ in place of α_n improves the convergence rate from $n^{-1/2}$ to n^{-2} .

It is also instructive to consider the convergence rate that arises when a deterministic grid is used to specify the points at which function evaluations are to be computed. In particular, suppose that the n function evaluations are performed at the points $1/(2n), 3/(2n), \dots, (2n-1)/(2n)$, and consider the estimator

$$\alpha_n^* = \frac{1}{n} \sum_{i=1}^n f\left(\frac{2i-1}{2n}\right).$$

Arguments similar to those used in the proof of Theorem 1 show that

$$n^2(\alpha_n^* - \alpha) \rightarrow (f^{(1)}(1) - f^{(1)}(0))/24. \quad (2.3)$$

Since $E(\chi_0^2) = E(\chi_1^2) = 2$, results (2.2) and (2.3) together suggest that the leading term of the expected error for the "randomized midpoint quadrature rule" is exactly twelve times that of the error for the deterministic midpoint quadrature rule. However, a feature of the estimator α_n^* is that it can not be recursively updated, since the grid for sample size $n+1$ is different from that for sample size n . In contrast, the estimator $\hat{\alpha}_{n+1}$ merely adds a function evaluation at the point U_{n+1} to those already computed at U_1, \dots, U_n . Thus, the factor of twelve can be viewed as a penalty paid to obtain an estimator that can be recursively updated. Of course, the recursive update of $\hat{\alpha}_{n+1}$ from $\hat{\alpha}_n$ would engender a significant computational burden. The integer i for which the subinterval $[L_{i-1}(n), L_i(n)]$ contains U_{n+1} would need to be determined, and the cost of finding i would increase with the sample size n . Indeed, for very large n , the cost could become prohibitively expensive.

This analysis in the one-dimensional setting suggests that use of function evaluation location information is particularly appropriate for situations where the proposed sample size n is moderate and function evaluations are expensive. Theorem 1 indicates that the slow convergence rate of $n^{-1/2}$ associated with naive Monte Carlo estimation is (for one-dimensional problems, at least) not a consequence of the irregular spacing of the function evaluations, but rather it results from the uniform weights that are used in constructing the sample mean. Limit (2.3) indicates that the rate of convergence for a properly weighted

Monte Carlo estimator is identical to that of the corresponding deterministic quadrature rule using a regularly spaced grid.

Another approach to improving the rate of convergence of α_n would be by developing a randomized version of Simpson's integration rule. The results observed here for quadrature rules suggest that the error of Simpson's rule on a regular deterministic grid would be asymptotically smaller (at least by a constant factor) than the error of a randomized version.

The primary purpose of this paper is to observe that the performance of Monte Carlo sampling schemes can sometimes be substantially improved by taking advantage of function evaluation location information. Given that deterministic quadrature rules are likely to be the methods of choice for one-dimensional integration problems (particularly when the integrand is smooth), the real impact of this observation lies in the possible use of location information for problems involving high-dimensional integrals, where Monte Carlo sampling is most frequently employed.

The theory presented here bears a superficial resemblance to a rotation sampling scheme proposed by Fishman and Huang (1983). Their method involves using a randomly translated regular grid with constant spacing, and it is effectively identical to the well known rectangular integration rule, as pointed out by Glynn and Whitt (1992). Our result, in contrast, involves function evaluations that lie on a randomly generated and irregularly spaced grid.

Conclusions

We have shown that the use of function evaluation location information can dramatically improve the performance of one-dimensional sampling schemes. This improvement arises via an appropriate reweighting of the function evaluations so that they no longer necessarily receive identical weights. The results obtained here suggest that it may be especially beneficial to judiciously incorporate location information when using Monte Carlo

sampling methods for estimation of high-dimensional integrals.¹

¹ The research of the first author was supported by National Science Foundation Contract No. DMS59204864. The research of the second author was supported by Army Research Office Contract No. DAAL03-91-G-0101 and by National Science Foundation Contract No. DDM91-01580.

Appendix

Proof of Theorem 1. Observe that

$$\alpha - \hat{\alpha}_n = \sum_{i=1}^n \int_{L_{i-1}(n)}^{L_i(n)} [f(x) - f(U_i(n))] dx. \quad (A.1)$$

Since f is twice continuously differentiable,

$$f(x) - f(U_i(n)) = f^{(1)}(U_i(n))(x - U_i(n)) + \frac{1}{2}f^{(2)}(\xi_z(n))(x - U_i(n))^2,$$

where $\xi_z(n)$ lies between x and $U_i(n)$. Now,

$$\sum_{i=2}^{n-1} \int_{L_{i-1}(n)}^{L_i(n)} f^{(1)}(U_i(n))(x - U_i(n)) dx = \frac{1}{2} \sum_{i=2}^{n-1} f^{(1)}(U_i(n))(\Delta_{i+1}(n)^2 - \Delta_i(n)^2),$$

where $\Delta_i(n) = (U_i(n) - U_{i-1}(n))/2$. By using summation by parts, the latter sum can be written as

$$\begin{aligned} & \frac{1}{2}f^{(1)}(U_{n-1}(n))\Delta_n(n)^2 - \frac{1}{2}f^{(1)}(U_1(n))\Delta_2(n)^2 \\ & - \frac{1}{2} \sum_{j=2}^{n-1} \{f^{(1)}(U_j(n)) - f^{(1)}(U_{j-1}(n))\} \Delta_j(n)^2 \\ & = \frac{1}{2}f^{(1)}(U_{n-1}(n))\Delta_n(n)^2 - \frac{1}{2}f^{(1)}(U_1(n))\Delta_2(n)^2 - \frac{1}{2} \sum_{j=2}^{n-1} f^{(2)}(\gamma_j(n))\Delta_j(n)^3, \end{aligned} \quad (A.2)$$

where $\gamma_j(n)$ lies between $U_{j-1}(n)$ and $U_j(n)$ ($j = 2, \dots, n-1$).

It is possible at this point to assume that $U_j(n) = T_j/T_{n+1}$ ($j = 1, \dots, n$), where T_j is the j th jump time of a Poisson process $N(t)$ having unit rate, since $(U_1(n), \dots, U_n(n))$ is distributed as $(T_1/T_{n+1}, \dots, T_n/T_{n+1})$ for each $n \geq 1$. (See, for example, Devroye 1982).

Since $f^{(2)}$ is continuous (and hence uniformly continuous) over $[0, 1]$, there exists an integer $k = k(\epsilon)$ for each $\epsilon > 0$ such that $|f^{(2)}(x) - f^{(2)}(y)| \leq \epsilon$ whenever $|x - y| \leq 1/k$.

Then

$$\begin{aligned} & \sum_{j=2}^{n-1} f^{(2)}(\gamma_j(n)) \Delta_j(n)^3 \\ & \geq \{f^{(2)}(0) - \epsilon\} \sum_{j=2}^{N(T_{n+1}/k)-1} \Delta_j(n)^3 + \sum_{i=2}^{k-1} \{f^{(2)}((i-1)/k) - \epsilon\} \sum_{j=N((i-1)T_{n+1}/k)}^{N(iT_{n+1}/k)-1} \Delta_j(n)^3 \\ & \quad + \{f^{(2)}((k-1)/k) - \epsilon\} \sum_{j=N((k-1)T_{n+1}/k)}^{n-1} \Delta_j(n)^3. \end{aligned}$$

Let $\tau_j = T_j - T_{j-1}$ ($j = 1, \dots, n+1$) and observe that $\tau_1, \dots, \tau_{n+1}$ are independent exponential random variables, each having mean 1. Since $T_n/n \rightarrow 1$ a.s. as $n \rightarrow \infty$, it is evident that

$$\begin{aligned} & n^2 \sum_{j=N((i-1)T_{n+1}/k)}^{N(iT_{n+1}/k)-1} \Delta_j(n)^3 \\ & = \left(\frac{n}{T_{n+1}}\right)^2 \frac{\sum_{j=N((i-1)T_{n+1}/k)}^{N(iT_{n+1}/k)-1} \tau_j^3}{N(iT_{n+1}/k) - N((i-1)T_{n+1}/k)} \frac{N(iT_{n+1}/k) - N((i-1)T_{n+1}/k)}{8T_{n+1}} \\ & \rightarrow \frac{1}{8k} E\tau_1^3 = \frac{3}{4k} \quad \text{a.s.} \end{aligned}$$

as $n \rightarrow \infty$. Hence,

$$\lim_{n \rightarrow \infty} n^2 \sum_{j=2}^{n-1} f^{(2)}(\gamma_j(n)) \Delta_j(n)^3 \geq \frac{3}{4k} \sum_{i=0}^{k-1} (f^{(2)}(i/k) - \epsilon) \quad \text{a.s.,}$$

and similarly,

$$\lim_{n \rightarrow \infty} n^2 \sum_{j=2}^{n-1} f^{(2)}(\gamma_j(n)) \Delta_j(n)^3 \leq \frac{3}{4k} \sum_{i=0}^{k-1} (f^{(2)}(i/k) + \epsilon) \quad \text{a.s.}$$

Since ϵ was arbitrary and since

$$\frac{1}{k} \sum_{i=0}^{k-1} f^{(2)}(i/k) \rightarrow \int_0^1 f^{(2)}(t) dt = f^{(1)}(1) - f^{(1)}(0),$$

it follows that

$$n^2 \sum_{j=2}^{n-1} f^{(2)}(\gamma_j(n)) \Delta_j(n)^3 \rightarrow \frac{3}{4} (f^{(1)}(1) - f^{(1)}(0)) \quad \text{a.s.} \quad (\text{A.3})$$

A similar argument proves that

$$\frac{n^2}{2} \sum_{i=2}^{n-1} \int_{L_{i-1}(n)}^{L_i(n)} f^{(2)}(\xi_x(n)) (x - U_i(n))^2 dx \rightarrow \frac{1}{4} (f^{(1)}(1) - f^{(1)}(0)) \quad \text{a.s.} \quad (\text{A.4})$$

as $n \rightarrow \infty$. Furthermore,

$$\int_0^{L_1(n)} f^{(1)}(U_1(n)) (x - U_1(n)) dx = f^{(1)}(U_1(n)) \left(\frac{\Delta_2(n)^2}{2} - \frac{U_1(n)^2}{2} \right)$$

and

$$n^2 f^{(1)}(U_1(n)) \frac{U_1(n)^2}{2} = n^2 f^{(1)} \left(\frac{T_1}{T_{n+1}} \right) \frac{\tau_1^2}{2T_{n+1}^2} \rightarrow f^{(1)}(0) \frac{\tau_1^2}{2} \quad \text{a.s.}$$

as $n \rightarrow \infty$. Whereas

$$\begin{aligned} \left| n^2 \int_0^{L_1(n)} f^{(2)}(\xi_x(n)) (x - U_1(n))^2 dx \right| &\leq n^2 \sup_{0 \leq x \leq 1} |f^{(2)}(x)| L_1(n)^3 \\ &\leq n^2 \sup_{0 \leq x \leq 1} |f^{(2)}(x)| \frac{T_2^3}{T_{n+1}^3} \rightarrow 0 \quad \text{a.s.,} \end{aligned}$$

it is evident that

$$n^2 \int_0^{L_1(n)} [f(x) - f(U_1(n))] dx - n^2 f^{(1)}(U_1(n)) \frac{\Delta_2(n)^2}{2} \rightarrow -f^{(1)}(0) \frac{\tau_1^2}{2} \quad \text{a.s.} \quad (\text{A.5})$$

as $n \rightarrow \infty$. A similar analysis holds for the error over the final subinterval $[L_{n-1}(n), 1]$.

Since the square of an exponential random variable with mean 1 has the chi-squared distribution with 2 degrees of freedom, Theorem 1 follows upon combining (A.1)-(A.5).

References

Davis, P. J. and Rabinowitz, P., *Methods of Numerical Integration*, Academic Press, New York, 1984.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE November 1992	3. REPORT TYPE AND DATES COVERED TEchnical		
4. TITLE AND SUBTITLE ON THE VALUE OF FUNCTION EVALUATION LOCATION INFORMATION IN MONTE CARLO SIMULATION		5. FUNDING NUMBERS DAAL03-91-G-0101		
6. AUTHOR(S) THOMAS J. DICICCIO AND PETER W. GLYNN				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Operations Research -4022 Terman Engineering Center Stanford University Stanford, CA 94305-4022		8. PERFORMING ORGANIZATION REPORT NUMBER 92-17		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709-2211		10. SPONSORING / MONITORING AGENCY REPORT NUMBER ARO 28809.11-MA-SW		
11. SUPPLEMENTARY NOTES The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.		12b. DISTRIBUTION CODE		
13. ABSTRACT (Maximum 200 words) The point estimator used in naive Monte Carlo sampling weights all the computed function evaluations equally, and it does not take into account the precise locations at which the function evaluations are made. In this note, we show for one-dimensional integration problems that if the weights are suitably modified to reflect the location information present in the sample, then the convergence rate of the Monte Carlo estimator can be dramatically improved from order $n^{-1/2}$ to order n^{-2} , where n is the number of function evaluations computed. 93 5 04 06 2				
14. SUBJECT TERMS Simulation, Monte Carlo methods, Numerical integration		15. NUMBER OF PAGES 10		
		16. PRICE CODE		
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

Devroye, L., "A log log law for maximal uniform spacings," *Ann. Probab.*, 10 (1982), 863-868.

Fishman, G. S. and B. D. Huang, "Antithetic variates revisited," *Comm. ACM*, 26 (1983), 964-971.

Glynn, P. W. and W. Whitt, "Efficiency of simulation estimators," *Oper. Res.*, 40 (1992), 505-520.